

An adaptive buddy check for observational quality control

DICK P. DEE^{1,2}
LEONID RUKHOVETS^{1,2}
RICARDO TODLING^{1,2}
ARLINDO M. DA SILVA²
JAY W. LARSON³

¹*General Sciences Corporation, Beltsville, Maryland*

²*Data Assimilation Office, NASA/Goddard Space Flight Center, Greenbelt, Maryland*

³*Mathematics and Computer Science Division, Argonne National Laboratory,
Argonne, Illinois*

Submitted to *Quart. J. Royal Meteor. Soc.*, September 8, 2000.

Abstract

An adaptive buddy check algorithm is presented that adjusts tolerances for outlier observations based on the variability of surrounding data. The algorithm derives from a statistical hypothesis test combined with maximum-likelihood covariance estimation. Its stability is shown to depend on the initial identification of outliers by a simple background check. The adaptive feature ensures that the final quality control decisions are not very sensitive to prescribed statistics of first-guess and observation errors, nor on other approximations introduced into the algorithm.

The implementation of the algorithm in a global atmospheric data assimilation is described. Its performance is contrasted with that of a non-adaptive buddy check, for the surface analysis of an extreme storm that took place in Europe on 27 December 1999. The adaptive algorithm allowed the inclusion of many important observations that differed greatly from the first guess and that would have been excluded on the basis of prescribed statistics. The analysis of the storm development was much improved as a result of these additional observations.

1 Introduction

This article describes a new algorithm for statistical quality control of observations for use in data assimilation. The method is based on the so-called buddy check, which assumes that the observables are spatially coherent, so that nearby measurements (buddies) should tend to confirm each other. If an outlier observation cannot be supported by its buddies, then it may well be corrupt, in which case it must be discarded. On the other hand, that outlier may contain genuine information about an unexpected event, in which case it should be used. Choosing one or the other of these two opposites can have a large impact on the assimilated product. The main feature of our new buddy check algorithm is that the rejection limits for outlier observations are adapted to the actual variability of surrounding data. This results in relatively tolerant acceptance criteria in synoptically active situations, and in more stringent criteria when conditions are quiet.

It is intuitively clear that some notion of reasonable differences among nearby observations is required in order to distinguish good data from bad. Such differences are due to the natural small-scale variability of the observables and to the inherent uncertainties of the measurement process. It is also clear that quality control is not a deterministic problem, in the sense that it is not generally possible to know with certainty whether an observation is good or bad. The performance of a quality test therefore must be measured in terms of probabilities. Ideally we would like a test for which the probability of failing a good observation is bounded by some small fraction (the significance level of the test), while the probability of failing a bad observation (the power of the test) is maximal. This is a classic problem in the theory of statistical hypothesis testing (Lehmann 1997), which, however, can be solved only if the probability distributions of both good and bad observations are known.

In a cycling data assimilation system, a short-term model forecast valid at or near the time of the observations is available as a first guess or background estimate for the observables. This can serve as prior information in a Bayesian formulation of the buddy check (Lorenc and Hammon 1988). In that case the probability distribution of the model forecast errors must be specified in addition to that of the observation errors. The probabilistic formulation of quality control is now well established in atmospheric data assimilation (Dharssi *et al.* 1992; Ingleby and Lorenc 1993). Variants of the buddy check have been implemented operationally in sequential statistical analysis schemes (Lorenc 1981, Woollen 1991) and more recently in the framework of variational assimilation (Andersson and Järvinen 1999). The buddy check is usually preceded by various sanity checks and other preliminary quality control procedures (e.g. Gandin 1988).

The main problem with the application of statistical quality control to atmospheric observations is the requirement for locally accurate information about

forecast and observation errors. Error covariance models used in current operational data assimilation systems do not accurately describe the dependence of errors on the flow. A quality control algorithm that relies on time- and space-averaged statistics tends to enforce those very statistics, simply by rejecting observations whenever the local variability is larger than usual. This can happen, for example, during the onset of an extreme weather event, when the forecast is poor and the error variances are underestimated. Any available observations are particularly important under those circumstances, so that the challenge is to design an effective, but not excessive, quality control algorithm that performs well in rapidly changing flow situations.

The outline of this paper is as follows. In Section 2 we present the derivation of an adaptive and iterative buddy check algorithm. Given an initial classification of the observations as either suspect or not, the buddy check can be regarded as a statistical test of the hypothesis that the observed discrepancies among the data are reasonable in view of their presumed probability distributions. The test can be made adaptive by re-estimating the required error statistics on the fly, using local non-suspect observations only. We prove that the algorithm is stable, in the sense that the adaptive tolerances are bounded if the initial identification of suspects is based on a simple background check. We then illustrate the performance of the algorithm, and the function of the adaptive feature in particular, by means of a simple contrived example.

In Section 3 we describe the implementation of the adaptive buddy check in the Goddard Earth Observing System Data Assimilation System (GEOS DAS). A background check, based on prescribed error statistics for the global analysis system, is used to define the initial set of suspects for the buddy check, but does not itself reject any observations. We briefly discuss the practical use of the background check in monitoring the validity of the prescribed error statistics. We then examine in detail the performance of the GEOS DAS quality control for the analysis of a severe storm that took place in Europe on 27 December 1999. The development of this storm was well observed but poorly analyzed by several operational centers. We show that the GEOS DAS quality control allowed many observations into the analysis that would have been excluded by a non-adaptive statistical algorithm. We also show that the final quality control decisions were largely insensitive to the prescribed error statistics.

2 The adaptive buddy check

Let the vector \mathbf{w}^o denote a subset of the observations which are subject to quality control. We will be flexible with regard to the specific composition of this subset, although we have in mind a mix of observations of different types, located within a limited spatial region. In any case, the starting point for the

buddy check is a preliminary classification of all elements of \mathbf{w}^o as either *suspect* or not. The observations that are not suspect are *buddies*. Observations may be flagged as suspect simply because they are outliers, or for any other reason. The buddy check tests the extent to which suspect observations are supported by their buddies. This is done by first predicting the values of the suspect data from the buddies, and then to test whether the discrepancy between the predictions and the actual observed values is reasonable or not.

The test can be formulated conveniently in terms of differences between the observations and some background estimate, which, in a cycling data assimilation system, is usually a short-term model forecast \mathbf{w}^f . The observed-minus-forecast residual vector \mathbf{v} is then defined by

$$\mathbf{v} = \mathbf{w}^o - \mathbf{h}(\mathbf{w}^f), \quad (1)$$

where \mathbf{h} is the observation operator associated with \mathbf{w}^o . In general, this operator involves nonlinear forward models relating observables (e.g., radiances) to model variables (e.g., temperatures). For direct observations of forecast model state variables \mathbf{h} is simply an interpolation from model grid points to observation locations. The residual \mathbf{v} is often referred to as the *innovation*, because it represents that part of the observational information which is not contained in the forecast.¹

The initial partitioning of all observations as either suspect or not is important, because only non-suspect observations are used in the prediction step of a buddy check. If some, but not all, suspect observations pass the check, then the partitioning should be updated accordingly and the buddy check should be repeated. This leads to an iterative procedure that terminates when no additional observations pass the test. The remaining suspect observations are then considered in gross error. We first describe a single iteration of the algorithm.

2.1 The optimal buddy check

The buddy check can be regarded as a statistical test of the assumption that all observations are devoid of gross errors. We introduce the null hypothesis

$$\mathbf{v} \sim \mathcal{N}(\mathbf{0}, \mathbf{S}), \quad (2)$$

or, in words, that all residuals are jointly normally distributed with zero mean and known covariance \mathbf{S} . We will later make allowance for the fact that this represents an idealization, even under the best of circumstances. For now we

¹This usage is imprecise and somewhat wishful in the context of data assimilation: observed-minus-forecast residuals are innovations only when the assimilation is optimal. See, for example, Anderson and Moore (1979, Section 5.3) for a correct definition of innovations.

assume that rejection of the null hypothesis implies that some of the data must be contaminated.

We partition the residual vector \mathbf{v} as

$$\mathbf{v} = \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix}, \quad (3)$$

where \mathbf{x} contains the residuals associated with suspect observations, and \mathbf{y} those associated with buddies. We then define corresponding blocks of the residual covariance,

$$\mathbf{S} = \begin{bmatrix} \mathbf{S}_x & \mathbf{S}_{xy} \\ \mathbf{S}_{xy}^T & \mathbf{S}_y \end{bmatrix}. \quad (4)$$

Under the null hypothesis, the conditional distribution of \mathbf{x} given \mathbf{y} is multivariate normal,

$$\mathbf{x}|\mathbf{y} \sim \mathcal{N}(\mathbf{x}^*, \mathbf{S}^*), \quad (5)$$

with

$$\mathbf{x}^* = \mathbf{S}_{xy} \mathbf{S}_y^{-1} \mathbf{y}, \quad (6)$$

$$\mathbf{S}^* = \mathbf{S}_x - \mathbf{S}_{xy} \mathbf{S}_y^{-1} \mathbf{S}_{xy}^T, \quad (7)$$

(Jazwinski 1970, Theorem 2.13). Each of these quantities is well-defined when \mathbf{S} is positive definite. In particular, $\mathbf{x}^* = \mathbf{x}^*(\mathbf{y})$ is the optimal estimate of \mathbf{x} based on \mathbf{y} (Jazwinski 1970, Theorem 5.3), and the matrix \mathbf{S}^* is the error covariance of this estimate. The main computation involved in (6) is the solution of the linear system $\mathbf{S}_y \mathbf{z} = \mathbf{y}$. Computation of (7) requires n additional solves of this system, where $n = \dim \mathbf{x}$ is the number of suspect observations.

With (5–7) in hand, we can apply rigorous statistical tests to the null hypothesis. The general idea is to compute the probability p that the suspect observations are consistent with the null hypothesis. If p is smaller than some prescribed threshold δ , we reject the hypothesis and conclude that it is likely that at least some of the data contain gross errors. The value of δ is called the *significance level* of the test.²

To be precise, (5–7) imply that the scalar quantity

$$\chi_n^2 = \frac{1}{n} (\mathbf{x} - \mathbf{x}^*)^T (\mathbf{S}^*)^{-1} (\mathbf{x} - \mathbf{x}^*) \quad (8)$$

²The significance level bounds the probability that the null hypothesis is falsely rejected, i.e. the probability of failing a good observation. Its value does not imply the probability of gross error, and it is therefore misleading to state that the null hypothesis may be rejected with confidence $1 - \delta$. See von Storch and Zwiers (1998, Chapter 4) for a lucid discussion of this and other subtleties associated with statistical hypothesis testing.

has a chi-squared distribution with n degrees of freedom (e.g., Tarantola 1987, Section 4.3.6). It follows that for any $\tau > 0$,

$$p = p(\chi_n^2 > \tau^2) = 1 - P\left(\frac{n}{2}, \frac{\tau^2}{2}\right), \quad (9)$$

where $P(a, x)$ is the incomplete gamma function defined by

$$P(a, x) = \frac{1}{\Gamma(a)} \int_x^\infty e^{-t} t^{a-1} dt. \quad (10)$$

Figure 1 shows the probability $p(\chi_n^2 > \tau^2)$ as a function of the number of suspect observations n , and of the tolerance parameter τ . The latter must be specified by the user. For example, we read from the figure that if $n = 16$ and the value of χ_n^2 is $6^2 = 36$, then we can reject the null hypothesis at the 1% significance level.

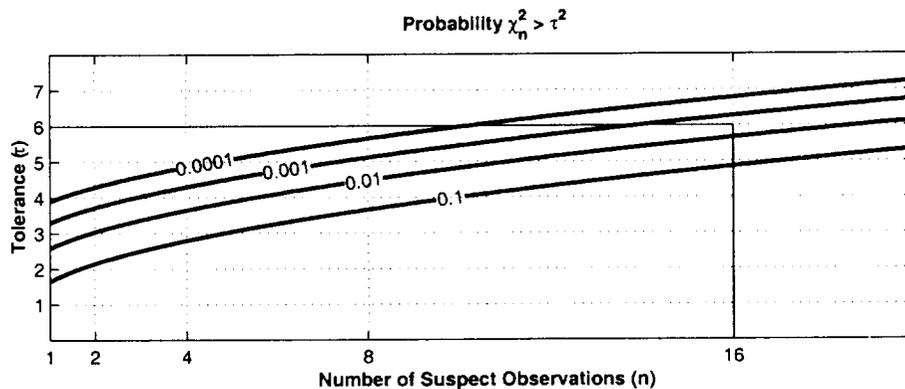


Figure 1: Chi-squared probabilities as a function of the number of suspect observations n , and of the tolerance parameter τ .

Once we conclude that one or more of the suspect data are likely to contain gross errors, we still need a procedure for marking individual observations for rejection. A simple approach is to consider the marginal distribution of each suspect residual element x_i in \mathbf{x} . All marginal distributions of a normal distribution are themselves normal (Jazwinski 1970, Theorem 2.12), which, together with (5-7) implies

$$x_i | \mathbf{y} \sim \mathcal{N}(x_i^*, S_{ii}^*), \quad (11)$$

with x_i^*, S_{ii}^* the appropriate elements in $\mathbf{x}^*, \mathbf{S}^*$, respectively. It follows that

$$p = p\left(\frac{|x_i - x_i^*|^2}{S_{ii}^*} > \tau^2\right) = \sqrt{\frac{2}{\pi}} \int_\tau^\infty e^{-t^2/2} dt. \quad (12)$$

For example, when $\tau = 3$ we have $p < 0.01$, which means that any observation for which $|x_i - x_i^*| > 3\sqrt{S_{ii}^*}$ can be rejected at the 1% significance level. Equation 12 corresponds to (9) with $n = 1$, so that these probabilities are also shown in Fig. 1.

In practice, the testing procedures and criteria for rejection should be designed to depend on the nature of the observations. For example, if \mathbf{x} corresponds to a simultaneous rawinsonde temperature and moisture report, then we might choose to reject both measurements if together they fail the chi-squared test. Similarly, if a height profile obtained from satellite data fails the test, then that profile should be rejected in its entirety. The advantage of simultaneously applying a single test to multiple data is that error correlations among the suspect residuals themselves can be properly taken into account. These and other possibilities are a matter of strategy and of practical viability, depending, for example, on available information about error covariances.

2.2 Adaptive tolerances

The procedure outlined so far presumes that rejection of the null hypothesis implies a strong likelihood of gross errors in the observations. A test of the null hypothesis, however, can also fail because of bad information about error characteristics. Clearly (2) is an idealization even in the absence of gross errors. For the buddy check to be effective, it must be reasonably robust with respect to any prescribed statistical parameters.

The specification of the covariance matrix \mathbf{S} in particular strongly influences the buddy check results. For example, (12) shows that the tolerance for a univariate test is proportional to the error variance S_{ii}^* , which, by (7), depends on \mathbf{S} . This means that the leniency of the optimal buddy check depends on the variability of the observed-minus-forecast residuals: the buddy check will less readily reject outlier observations if the deviations are expected to be large. This behavior is clearly desirable, but it crucially depends on the ability to specify locally accurate covariances for the data residuals.

We can show from (1) (e.g. Dee 1995) that the covariance matrix \mathbf{S} of the residual vector \mathbf{v} is

$$\mathbf{S} \approx \mathbf{R} + \mathbf{H}\mathbf{P}^f\mathbf{H}^T, \quad (13)$$

where \mathbf{R} is the observation error covariance, \mathbf{H} is the linearized observation operator defined by

$$\mathbf{H} = \left. \frac{\partial \mathbf{h}}{\partial \mathbf{w}} \right|_{\mathbf{w}=\mathbf{w}^f}, \quad (14)$$

and \mathbf{P}^f is the forecast error covariance. Equation (13) would be exact if forecast and observation errors were statistically independent and if the observation operator were linear. In practice, error covariances in operational data assimilation systems are difficult to estimate and may be accurate in a space- and time-average sense only.

The prescription of \mathbf{S} inherited from the global analysis system therefore represents, at best, a reasonable first guess of the residual covariance matrix based on time- and space-averaged statistics. The buddy check, however, is inherently local in nature, and we should therefore attempt to make adjustments to the tolerances. We can do so by introducing a parameter α to rescale the prescribed covariances. Replacing the the null hypothesis (2) by

$$\mathbf{v} \sim \mathcal{N}(\mathbf{0}, \alpha^2 \mathbf{S}), \quad (15)$$

the maximum-likelihood estimate of α based on the (non-suspect) data residuals \mathbf{y} is given by

$$\alpha^2 = \frac{1}{m} \mathbf{y}^T \mathbf{S}_y^{-1} \mathbf{y}, \quad (16)$$

where $m = \dim \mathbf{y}$ is the number of buddies (Dee 1995, Section 4). This computation is practically cost-free, by virtue of (6). Equation (7) shows that to rescale the residual covariance \mathbf{S} is equivalent to rescaling the conditional covariance \mathbf{S}^* . This in turn has the effect of modifying the tolerance for the buddy check in (12), replacing τ by $\alpha\tau$.

A uniform rescaling of the residual covariances as in (15) is reasonable only if the observations are associated with a limited region in space. To alleviate this restriction one could generalize (15) by introducing additional parameters (Dee 1995), but this would complicate the algorithm. A practical implementation must therefore include a strategy for subdividing the observations into suitable subsets, such that an adaptive buddy check with uniform rescaling can be applied to each subset separately.

Adapting tolerances on the fly based on current observations requires some care, to ensure that the algorithm remains stable. Equation (16) implies

$$\frac{(\mathbf{y}^T \mathbf{y})/m}{\lambda_{max}(\mathbf{S}_y)} < \alpha^2 < \frac{(\mathbf{y}^T \mathbf{y})/m}{\lambda_{min}(\mathbf{S}_y)}, \quad (17)$$

where $\lambda_{min}(\mathbf{S}_y)$ and $\lambda_{max}(\mathbf{S}_y)$ denote the smallest and largest eigenvalues of \mathbf{S}_y , respectively. Thus, the inclusion in \mathbf{y} of a few extreme outliers can cause excessive amplification of the residual covariances, resulting in a buddy check which is unable to detect gross errors. This can be prevented by making sure that all non-suspect residuals are initially bounded, as in

$$y_i^2 < \tau_b^2 S_{y,ii} \quad \text{for } i = 1, \dots, m, \quad (18)$$

where τ_b is a prescribed tolerance parameter. This represents a simple *background check*, strictly based on prescribed statistics. We then have

$$\alpha^2 < \tau_b^2 \frac{(\text{trace } \mathbf{S}_y)/m}{\lambda_{\min}(\mathbf{S}_y)}, \quad (19)$$

which gives an upper bound for the rescaling parameter that depends on the prescribed covariances only. Note that λ_{\min} can be estimated based on (13); for example, in case of independent observations with prescribed error standard deviation σ^o , we have $0 < (\sigma^o)^2 < \lambda_{\min}$.

The variance of the maximum-likelihood estimate α given by (16) is proportional to m^{-1} for sufficiently large number of buddies m (Dee 1995, Section 4). However, the estimate is not meaningful when the number of buddies m is very small. To account for this situation, we can introduce a smoothing parameter $m^* \geq 0$ and use instead

$$\alpha^2 = \frac{\mathbf{y}^T \mathbf{S}_y^{-1} \mathbf{y} + m^*}{m + m^*}. \quad (20)$$

This has the effect of reverting to the prescribed covariance \mathbf{S} when $m \ll m^*$.

2.3 Summary of the algorithm

The following algorithm implements a simple background check, followed by an iterative, adaptive buddy check, applied to a subset \mathbf{v} of observed-minus-forecast residuals with prescribed covariance \mathbf{S} . The background check serves to define the initial set $\mathbf{x}^{(0)}$ of suspects, as well as to bound the non-suspect residuals used for the buddy check. Upon completion, any residuals that remain in $\mathbf{x}^{(k)}$ are considered in gross error.

$$\mathbf{x}^{(0)} = \{v_i \in \mathbf{v} \text{ such that } |v_i| > \tau_b \sqrt{S_{ii}}\} \quad (\text{A.0})$$

for $k = 1, 2, \dots$

$$\mathbf{y} = \{v_i \in \mathbf{v} \text{ such that } v_i \notin \mathbf{x}^{(k-1)}\} \quad (\text{A.1})$$

$$\mathbf{x}^* = \mathbf{S}_{xy} \mathbf{S}_y^{-1} \mathbf{y} \quad (\text{A.2})$$

$$\mathbf{S}^* = \mathbf{S}_x - \mathbf{S}_{xy} \mathbf{S}_y^{-1} \mathbf{S}_{xy}^T \quad (\text{A.3})$$

$$\alpha^2 = \frac{\mathbf{y}^T \mathbf{S}_y^{-1} \mathbf{y} + m^*}{\dim \mathbf{y} + m^*} \quad (\text{A.4})$$

$$\mathbf{x}^{(k)} = \{x_i \in \mathbf{x}^{(k-1)} \text{ such that } |x_i - x_i^*| > \alpha \tau \sqrt{S_{ii}^*}\} \quad (\text{A.5})$$

$$\text{if } \dim \mathbf{x}^{(k)} = \dim \mathbf{x}^{(k-1)} \text{ or } \dim \mathbf{x}^{(k)} = 0 \text{ then stop} \quad (\text{A.6})$$

end

Since each iteration results in a decrease in the number of suspects by at least one, the algorithm is guaranteed to converge in at most $n = \dim \mathbf{x}^{(0)}$ iterations. In practice, convergence is much faster. The user must specify the tolerance parameters τ_b , τ and the relaxation parameter m^* . We usually take $\tau_b = 2$ for the background check; equation (12) then predicts that roughly 4.5% of all residuals should be initially marked as suspect. For the buddy check tolerances we take $\tau = 3$, corresponding to a significance level of about 0.3%. With $m^* = 0$ the algorithm fully adjusts the tolerances for the buddy check during each iteration, based on the current set of non-suspects. Setting $m^* \gg m$ effectively turns off the adaptive feature in the algorithm, which then completely relies on prescribed covariances.

2.4 A simple illustration

Our experiment simulates a one-dimensional domain with 32 equally spaced observation locations, labeled $i = 1, 2, \dots, 32$. The analysis system operates under the null hypothesis

$$\mathbf{v} \sim \mathcal{N}(\mathbf{0}, \mathbf{S}), \quad (22)$$

with

$$S_{ij} = \begin{cases} 1, & \text{for } i = j, \\ \frac{1}{2}e^{-0.2(i-j)^2}, & \text{otherwise.} \end{cases} \quad (23)$$

This simple model represents residuals with a spatially uncorrelated observation error component and a correlated forecast error component; see Dee and da Silva 1999.

The actual residuals, however, are distributed according to

$$\mathbf{v} \sim \mathcal{N}(\mathbf{b}, \sigma^2 \mathbf{S}), \quad (24)$$

with

$$b_j = 2 \sin \frac{\pi j}{32}, \quad (25)$$

$$\sigma = 2. \quad (26)$$

Here \mathbf{b} represents a bias in the residual, and σ is a noise amplification factor. Both \mathbf{b} and σ are unknown to the algorithm. This type of situation can easily arise at a time when forecast errors are spatially coherent and larger than usual in some region. The challenge for a buddy check is then to recognize that the data do not contain any gross errors, even though many of the residuals may be much larger than expected.

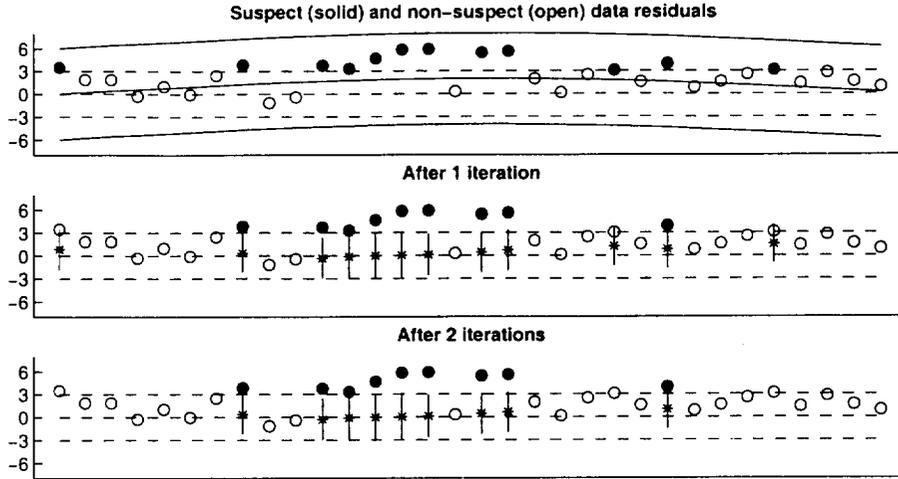


Figure 2: Illustration of the performance of the iterated buddy check with prescribed tolerances, for the contrived example described in the text. The solid curves in the top panel show the actual expected range for the residuals. The dashed curves indicate the tolerances used by the algorithm, which are based on prescribed statistics.

Figure 2 illustrates what happens when the non-adaptive algorithm ($m^* \gg m$) is applied. The circles in the top panel mark a single realization of (24) produced with a random number generator. The solid curves indicate the actual unconditional expected range $b_j \pm \tau\sigma$ for the vast majority (about 99.7%) of residuals. The dashed horizontal lines show the unconditional range $0 \pm \tau$ anticipated by the analysis system; all residuals that are outside this range (indicated by solid circles in the figure) are initially considered suspect.

The center panel of Fig. 2 summarizes the situation at the end of the first iteration of the algorithm. For each suspect residual x_i , an asterisk marks its conditional expectation x_i^* given all non-suspect data. The vertical bars represent the conditional range $x_i^* \pm \tau\sqrt{S_{ii}^*}$. Notice that this range is always slightly smaller than the unconditional range—more so when there are buddies nearby. Three residuals are found to lie within the conditional range, and are therefore no longer considered suspect. After updating the set of buddies, a second iteration of the algorithm does not change the status of any of the remaining suspects. The algorithm then terminates. All observations marked by solid circles in the bottom panel of Fig. 2 end up failing the buddy check.

Figure 3 summarizes the performance of the adaptive algorithm ($m^* = 0$), for the same set of residuals and using the same prescribed covariance information. The second panel shows that, already after the first iteration, more of the sus-

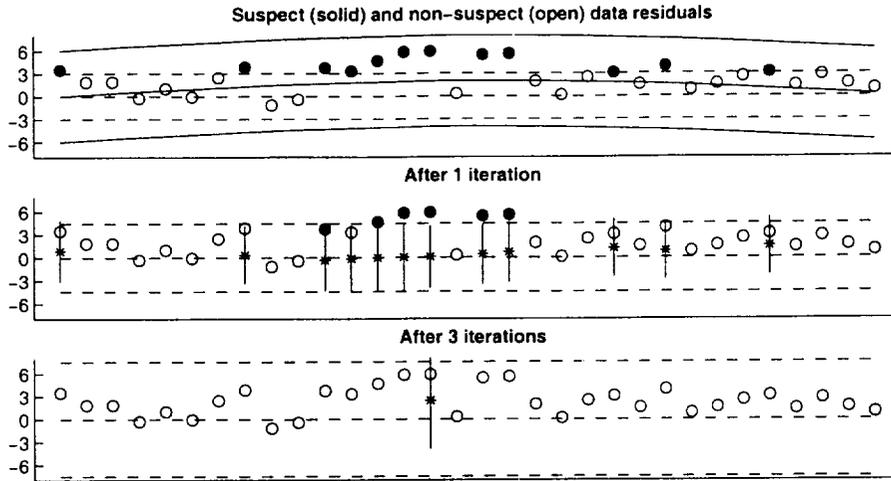


Figure 3: As Fig. 2, but for the adaptive buddy check. The tolerances (indicated by the dashed curves) expand during the iterations to accommodate the actual variability of the data.

pect residuals lie within the conditional range. The reason is that the algorithm senses that the prescribed bounds are too conservative, based on the variability of the initial non-suspects. The dashed horizontal lines now indicate the adjusted range $0 \pm \alpha\tau$, which has expanded slightly. Subsequent iterations allow the remaining suspect residuals to be confirmed by the growing set of buddies. The bottom panel of the figure shows that, in fact, all residuals end up passing the buddy check. Note that any residual lying outside the final adapted range (indicated by the dashed lines in the bottom panel) would have been rejected.

A critical safety feature in the adaptive algorithm is that all tolerance adjustments are based on non-suspect data only. Therefore, any outlier observations that are not supported by their neighbors will be rejected if they are initially flagged as suspect. This condition on the initial partitioning is guaranteed by the first step of the algorithm, which is a simple background check based on prescribed statistics only. In the pathological case when *all* residuals fail the background check the algorithm will reject them all. This situation can occur in practice when an extreme event is observed by a small number of isolated observations.

3 Implementation in GEOS DAS

The GEOS DAS Version 3 produces global atmospheric data sets at 6-hourly intervals on a $1^\circ \times 1^\circ$ latitude-longitude grid and on 48 vertical levels. The core of the system consists of an atmospheric GCM (Takacs and Suarez 1996), a physical-space statistical analysis system (PSAS) (Cohn *et al.* 1998), the statistical quality control (SQC) described here, and various interface functions. Apart from conventional observations, the system accepts geopotential heights retrieved from TIROS operational vertical sounder (TOVS) data, cloud-drift wind retrievals, and SSM/I surface winds and total precipitable water. The final, assimilated data products are obtained from the analyzed fields by means of the incremental analysis update (IAU) procedure (Bloom *et al.* 1996).

The SQC is invoked just prior to computing the global analysis, and therefore represents the final line of defense against the inclusion of bad data. All observations have passed various sanity checks and other preliminary quality control procedures by the time they are presented to the SQC. Some may be marked suspect for various reasons. Input for the SQC consists of observed-minus-forecast residuals and a preliminary estimate of their variances, derived from prescribed error statistics for the global analysis system. Observation error standard deviations for most GEOS DAS data types were estimated using maximum-likelihood techniques (Dee and da Silva 1999; Dee *et al.* 1999). Forecast error standard deviations currently used in GEOS DAS are global, spatially variable estimates based on monthly statistics of rawinsonde and TOVS observed-minus-forecast height residuals (DAO 1996).

3.1 The background check

The SQC first performs a simple background check of all observations against the 6-hour forecast, as in (A.0). This test does not actually reject any observations, but marks as suspect all residuals v_i for which

$$v_i^2 > \tau_b^2 [(\sigma_i^o)^2 + (\sigma_i^f)^2]. \quad (27)$$

Here τ_b is a tolerance parameter and σ_i^o, σ_i^f are prescribed observation and forecast error standard deviations, respectively, appropriate for the observation type and location.

The rate at which the background check flags data provides a useful consistency check on the prescribed statistics. Equation (12) predicts that normal, independently distributed residuals should fail the background check at an average rate of about 4.5% when $\tau_b = 2$, which is the value currently used in GEOS DAS. Actual rates will vary even with correctly specified statistics, because residuals are neither normal nor independently distributed. Consistently large deviations

Source	Type	Data count	Background check	Buddy check
Rawinsondes	(z)	466 903	6.07	1.25
	(u, v)	822 632	6.34	1.90
	(q)	197 582	6.83	1.15
TOVS	(z)	6 271 375	2.93	0.61
Aircraft	(u, v)	638 136	8.18	1.67
Pilot balloons	(u, v)	147 172	5.56	2.02
Cloud drift	(u, v)	1 061 248	4.03	1.07
Surface stations	(p_{sl})	790 857	9.41	1.22
Ships	(p_{sl})	131 520	4.78	0.88
	(u, v)	238 186	5.85	2.77
Buoys	(p_{sl})	203 705	5.38	1.60
	(u, v)	186 278	5.10	1.84
SSM/I	(u, v)	4 638 572	0.27	0.04

Table 1: Summary of GEOS DAS Statistical Quality Control monitoring for January 2000. Background and buddy check rates are in percents.

from the predicted rate suggest a problem with either the prescribed error standard deviations or with other assumptions about errors incorporated into the analysis system. Intermittently large deviations may indicate transitional problems with the DAS, such as an exceptionally poor forecast or a breakdown of the observing system. Background check results should be monitored for each observing system and data type, broken down by region and vertical level.

Table 1 summarizes the SQC monitoring for GEOS DAS during January 2000. The symbols $z, u, v, q,$ and p_{sl} in the second column stand for the analyzed quantities geopotential height, zonal wind, meridional wind, water vapor mixing ratio, and sea-level pressure, respectively. The first numeric column shows the number of scalar observations presented to the SQC for each data type (counting each wind vector report as 2 observations). The second numeric column shows the background check rates, in percent. The final column contains the rejection rates for the buddy check, which we discuss in the next section.

None of the actual background check rates shown in Table 1 exactly match the ideal rate. Sea-level pressure data from surface stations are flagged at a higher rate because the distribution of residuals for that data type is slightly peaked, with thick tails relative to the normal distribution. The tails primarily result from the use of (extrapolated) sea-level pressure data over topography. The low

rate for TOVS height retrievals appears to be due to an earlier removal of outliers during the retrieval process. It may also result from a statistical dependence between retrieval errors and forecast errors, as well as underestimation by the analysis system of stratospheric forecast errors. The exceedingly low rate for SSM/I winds indicates that the prescribed observation error standard deviations for this data type are too high and need to be adjusted.

3.2 The buddy check

Following the background check, the SQC applies a buddy check to decide the final status of all observations. Each suspect residual is tested against nearby non-suspect residuals, using an algorithm essentially as described in Section 2. Wind vector data are excluded when either of the components fail the buddy check. Similarly, an entire TOVS height profile is excluded if any single height residual from that profile fails the buddy check. The final column in Table 1 shows the average rejection rates for each of the GEOS DAS data types during January 2000.

We introduced several approximations in the current implementation of the algorithm in the SQC. The main simplification is that the buddy check is univariate, in the sense that only data of the same type (but possibly from different instruments) are used to test each suspect observation. In a multivariate check, confirmation of an extremely low sea-level pressure observation, for example, might be found in nearby cyclonic wind observations. Furthermore, the local analysis performed in each buddy check uses a single iteration of the successive corrections method (Daley 1991, Chapter 3) rather than a statistical interpolation. It would be more elegant to call the analysis component of the DAS to solve (A.2) in each instance, but that is currently not practical.

In its current configuration, the computational cost of the SQC represents about 6% of the total cost of a global analysis. The main portion of this is expended during the first iteration of the buddy check, which is applied to all observations that are initially flagged as suspect. The majority (typically 85-90%) are reaccepted after the first iteration. The cost can be significantly reduced, if necessary, by increasing the tolerance parameter for the background check, which would result in a smaller pool of initial suspects. Our current choice $\tau_b = 2$ is conservative and could probably be increased without changing the final result of the buddy check after convergence. The total number of iterations of the buddy check needed for convergence is typically between 3 and 6.

In spite of the approximations, the buddy check as implemented in the SQC retains the main features of the algorithm described in Section 2: it is based on a local analysis of nearby data which is both adaptive and iterative. In analogy with the simple contrived example presented in Section 2.4, we illustrate the

performance of the SQC using the following case study.

3.3 Storm of 27 December 1999

Two severe storms hit Europe in succession at 06 UTC on 26 December 1999 and at 18 UTC on 27 December 1999, causing significant damage and a great deal of media attention. Both storms were poorly predicted by many weather services, in spite of the fact that they were well observed by a variety of observing systems. Preliminary indications are that the forecast problems were largely due to inadequate data assimilation procedures, since many models were able to predict the storms several days ahead but lost track of them in subsequent short-range forecasts (P. Undén, *pers. comm.*; see also the *Official SRNWP/EUCOS Report on the December 1999 Storms*, available on the internet at <http://srnwp.sma.ch/workshops/FinalReport.html>). Because of background error covariance specifications that are inappropriate for extreme situations, the available observations were interpreted incorrectly, and, in some cases, were excluded from the analysis altogether. At Météo France the medium-range forecasts of the storms were reasonably good, but a large number of crucial surface observations were not assimilated when the second storm hit the coast. These observations were excluded by a simple statistical background check because the model first-guess became very poor (J.-N. Thépaut, *pers. comm.*).

To illustrate this problem, we show in Figure 4 the result of an experiment with GEOS DAS in which the adaptive feature of the buddy check is switched off. The three panels in the figure show the development of the second storm in a sequence of sea-level pressure analyses. Each panel contains the locations of all available pressure observations that took place within a 6-hour window centered at the analysis time (some locations correspond to multiple observations). The color green indicates that the observations at that location passed the background check, yellow means that at least one observation failed the background check but subsequently passed the buddy check, and red means that at least one observation failed the buddy check and therefore did not enter the analysis. Not shown are the variety of near-surface wind observations that were analyzed as well, originating from ships, buoys, and SSM/I retrievals. Both the background check and the buddy check in this case are strictly based on prescribed statistics.

At 12 UTC and at 18 UTC the background check flagged a large number of observations, at a rate much higher than usual because of the poor 6-hour forecast. A fair number of these ultimately did pass the buddy check, but the most crucial observations in the vicinity of the depression were rejected. The analysis at 12 UTC shows a weak low of 989.3 hPa located over the Celtic Sea. The actual low is further to the south-west, in the vicinity of 8W, 48N, where an entire cluster of observations was rejected by the buddy check. The lowest three in this cluster averaged 974.2 hPa, which is about 17 hPa below the analyzed

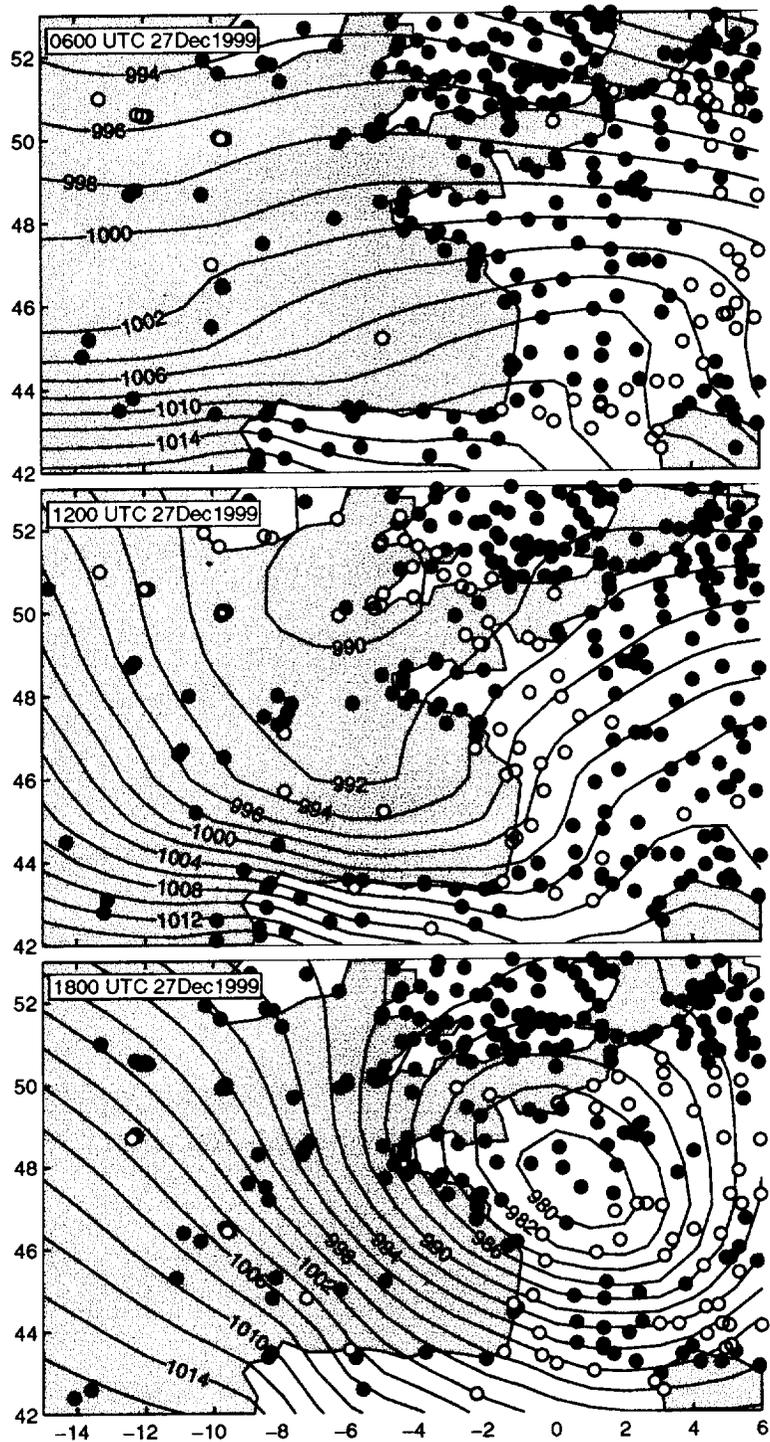


Figure 4: Successive GEOS surface analyses showing the development of the 27 December 1999 storm, obtained when using a buddy check whose tolerances are based on prescribed statistics. Red disks mark the locations of sea-level pressure observations that were rejected by the buddy check, yellow disks correspond to observations that failed the background check but subsequently passed the buddy check, and green disks correspond to those that passed the background check.

minimum. At 18 UTC the depression over France is too weak by about 10 hPa; the analyzed low there is 978.8 hPa, while the three lowest observations at that time averaged only 968.8 hPa.

Figure 5 shows the GEOS DAS analyses obtained by using the adaptive buddy check. Almost all of the observations that were flagged by the background check were ultimately allowed into the analysis. The definition of the storm is now much better. The analyzed low at the height of the storm is 973.6 hPa, which is 5.2 hPa deeper as a result of the additional observations. The first-guess low was 980.2 hPa, about 4.5 hPa deeper due to the improved initial conditions at 12 UTC. Although the model still has difficulty capturing the storm in its full strength, the inclusion of the available observations in this case clearly improves the assimilation.

We now take a close look at the observations that failed the buddy check. The single rejected observation at 12 UTC, located near the center of the depression, is a ship report of 879.8 hPa, which is clearly in error. On the other hand, for the 18 UTC analysis a report of 971.3 hPa from a surface station on the French coast was rejected, even though it was probably accurate. Three successive reports were issued from that station, valid at 17 UTC (971.3 hPa), 18 UTC (976.8 hPa), and 19 UTC (982.8 hPa). These values are consistent with a rapid inland movement of the storm. The first report was rejected by the buddy check because it differed greatly from the first-guess value of 983.6 hPa and could not be confirmed by surrounding observations (including the two later reports from the same station).

For similar reasons, seven sea-level pressure ship reports were excluded from the 6 UTC analysis, all of which were probably accurate. These observations, whose locations are marked by the four red disks in the top panel of Figure 5, all took place toward the end of the 6-hour analysis window and showed a drop in sea-level pressure of about 10hPa compared to reports just a few hours earlier. In addition, there were some late wind reports from nearby buoys (not shown) that indicated a change to easterly winds associated with a developing cyclonic circulation. A multivariate buddy check might have been able to take advantage of this information. If these data had been included in the analysis, the model would perhaps have been able to detect the developing depression at an earlier stage.

The underlying problem here is that it is not possible to accurately represent rapidly moving storm systems when all observations within a 6-hour time window are treated as if they occurred simultaneously. This is clearly a shortcoming of the assimilation system. Work is well underway to reduce the length of the analysis window in GEOS DAS.

The most important practical advantage of the adaptive buddy check is that the final quality control decisions are robust with respect to the prescribed

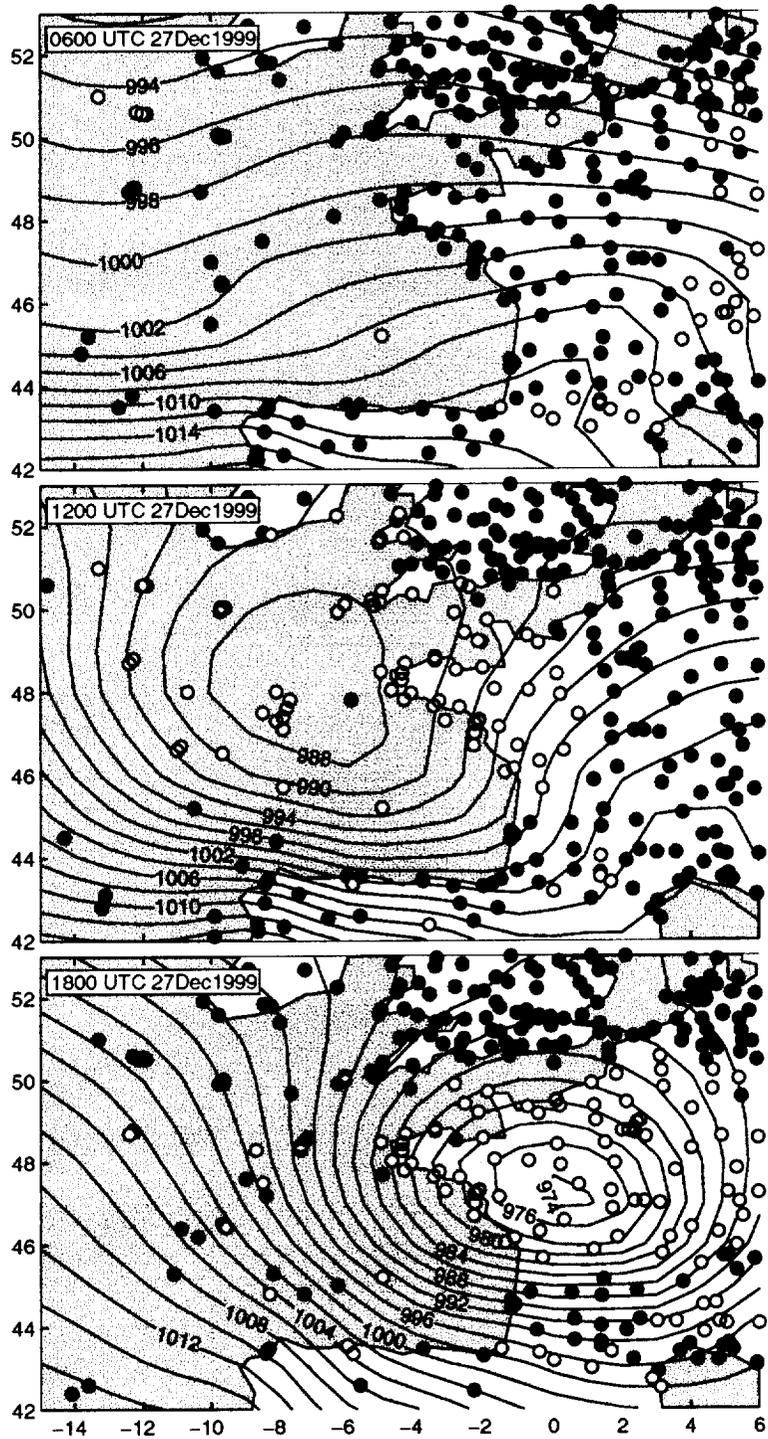


Figure 5: As Fig. 4, but using an adaptive buddy check.

forecast and observation error statistics. To demonstrate this we changed the tolerance parameter for the background check, which has the effect of modifying the initial pool of suspects. The case just discussed uses $\tau_b = 2$, which results in 57 suspects at 6 UTC, 153 at 12 UTC, and 186 at 18 UTC. We repeated the quality control with $\tau_b = 3$, which reduced the number of initial suspects to 10, 84, and 111, respectively. However, the final result of the buddy check was unchanged for all observations.

4 Conclusion

We have presented an adaptive buddy check algorithm that locally adjusts the tolerances for outlier observations, on the basis of the variability of the prevailing flow. This adaptive feature ensures that the final quality control decisions depend primarily on the surrounding observations and only weakly on prescribed error statistics. This is an important practical advantage, since the error statistics tend to be least accurate in situations that are difficult to forecast, which is precisely when quality control decisions have their largest impact. Prescribed statistics must still be used for the initial identification of outlier observations in order to ensure the stability of the adaptive algorithm.

We illustrated the performance of the algorithm using two analogous examples. The first of these used contrived observations, drawn from an error distribution which is very different from that initially presumed. The second example was a realistic case study based on the GEOS DAS analysis of an extreme storm event. In both examples we first applied an iterative buddy check with fixed tolerances based on prescribed error statistics, which led to the exclusion of many important observations. We then applied the adaptive algorithm, which was able to adjust the tolerances and bring in the great majority of the observations. For the GEOS DAS case study we showed that the analysis of the storm was greatly improved as a result.

The performance of the adaptive buddy check has been monitored since late 1998 in successive versions of the GEOS DAS, both in terms of data counts and in individual case studies. We consistently find that the final quality control decisions are not very sensitive to the various approximations that have been introduced in the implementation of the algorithm. The main sensitivity, which is to the prescribed statistics, has been largely removed by making the buddy check adaptive. Strict adherence to the statistical theory is clearly less important in this instance than a practical implementation that takes account of the largest uncertainties in the problem at hand.

Acknowledgements. Thanks to Gerard Cats for first suggesting this application of adaptive error estimation, to Jim Stobie for constructive criticism, and to

Jean-Noël Thépaut for pointing to the role of quality control in the analysis of the 27 December 1999 storm.

References

- Andersson, B. D. O., and J. B. Moore, 1979: *Optimal Filtering*. Prentice-Hall, Englewood Cliffs, 357pp.
- Andersson, E., and H. Järvinen, 1999: Variational quality control. *Quart. J. Royal Meteor. Soc.*, **125**, 697–722.
- Bloom, S. C., L. L. Takacs, A. M. da Silva, and D. Ledvina, 1996: Data assimilation using incremental analysis updates. *Mon. Wea. Rev.*, **124**, 1256–1271.
- Cohn, S. E., A. da Silva, J. Guo, M. Sienkiewicz, and D. Lamich, 1998: Assessing the effects of data selection with the DAO physical-space statistical analysis system. *Mon. Wea. Rev.*, **126**, 2913–2926.
- Daley, R., 1991: *Atmospheric Data Analysis*. Cambridge University Press, Cambridge, 457pp.
- DAO, 1996: *Algorithm Theoretical Basis Document Version 1.01*. Data Assimilation Office, NASA/Goddard Space Flight Center, Greenbelt, MD 20771.
- Dee, D. P., 1995: On-line estimation of error covariance parameters for atmospheric data assimilation. *Mon. Wea. Rev.*, **123**, 1128–1145.
- Dee, D. P., and A. M. da Silva, 1999: Maximum-likelihood estimation of forecast and observation error covariance parameters. Part I: Methodology. *Mon. Wea. Rev.*, **124**, 1822–1834.
- Dee, D. P., G. Gaspari, C. Redder, L. Rukhovets, and A. M. da Silva, 1999: Maximum-likelihood estimation of forecast and observation error covariance parameters. Part II: Applications. *Mon. Wea. Rev.*, **124**, 1835–1849.
- Dharssi, I., A. C. Lorenc, and N. B. Ingleby, 1992: Treatment of gross errors using maximum probability theory. *Quart. J. Royal Meteor. Soc.*, **118**, 1017–1036.
- Gandin, L. S., 1988: Complex quality control of meteorological observations. *Mon. Wea. Rev.*, **116**, 1137–1156.
- Ingleby, N. B., and A. C. Lorenc, 1993: Bayesian quality control using multivariate normal distributions. *Quart. J. Royal Meteor. Soc.*, **119**, 1195–1225.

- Jazwinski, A. H., 1970: *Stochastic Processes and Filtering Theory*, Academic Press, New York, 376pp.
- Lehmann, E. L., 1997: *Testing Statistical Hypotheses*, Springer, New York, 600pp.
- Lorenc, A. C., 1981: A global three-dimensional multivariate statistical interpolation scheme. *Mon. Wea. Rev.*, **109**, 701–721.
- Lorenc, A. C., and O. Hammon, 1988: Objective quality control of observations using Bayesian methods: Theory, and a practical implementation. *Quart. J. Royal Meteor. Soc.*, **114**, 515–543.
- von Storch, H., and F. W. Zwiers, 1998: *Statistical Analysis in Climate Research*, Cambridge University Press, Cambridge, 484pp.
- Takacs, L. L., and M. J. Suarez, 1996: Dynamical aspects of climate simulations using the GEOS general circulation model. *NASA Technical Memorandum 104606*, **10**, 56pp.
- Tarantola, A., 1987: *Inverse Problem Theory*, Elsevier, Amsterdam, 613pp.
- Woollen, J. S., 1991: New NMC operational OI quality control. Pp. 24–27 in: *Proc. 9th Conf. Numerical Weather Prediction*, Denver, American Meteorological Society.